CLAIMS

WHAT IS CLAIMED IS:

1. A text mining method for extracting features of documents using a term-document matrix consisting of vectors corresponding to index terms representing the contents of the documents, wherein contributions of the index terms act on respective elements of the term-document matrix, said method comprising:

a basis vector calculating step of calculating a basis vector spanning a feature space, in which mutually associated documents and terms are located in proximity with each other, based on a steepest descent method minimizing a cost;

a feature extracting step of calculating a parameter for normalizing the features using the term-document matrix and the basis vector and extracting the features on the basis of the parameter; and

a term-document matrix updating step of updating the term-document matrix to a difference between the term-document matrix, to which the basis vector is not applied, and the term-document matrix, to which the basis vector is applied.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

2. A text mining method for extracting features of documents as claimed in claim 1, wherein the cost is defined as a second-order cost of the difference between the term-document matrix, to which the basis vector is not applied, and the term-document matrix, to which the basis vector is applied.

3. A text mining method for extracting features of documents as claimed in claim 2, wherein said basis vector calculating step comprises:

an initializing step of initializing a value of the basis vector;

a basis vector updating step of updating the value of the basis vector;

a variation degree calculating step of calculating a variation degree of the value of the basis vector;

a judging step of making a judgment whether a repetition process is to be terminated or not using the variation degree of the basis vector; and

a counting step of counting the number of times of said repetition process.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

4. A text mining method for extracting features of documents as claimed in claim 3, wherein said basis vector updating step updates the basis vector using a current value of the basis vector, the term-document matrix and an updating ratio controlling the updating degree of the basis vector.

5. A text mining method for extracting features of documents as claimed in claim 4, wherein, when all basis vectors and normalizing parameters required in extracting the features have been already obtained, the calculation of normalizing parameters in said basis vector calculating step and the execution of said feature extracting step are omitted, and said feature extracting step extracts the features using the basis vectors and the normalizing parameters that have been already obtained.

6. A text mining method for extracting features of documents as claimed in claim 3, wherein, when all basis vectors and normalizing parameters required in extracting the features have been already obtained, the calculation of normalizing parameters in said basis vector calculating step and the execution of said feature extracting step are omitted, and said feature extracting step extracts the features using the basis vectors and the normalizing parameters that have been already obtained.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

7. A text mining method for extracting features of documents as claimed in claim 2, wherein, when all basis vectors and normalizing parameters required in extracting the features have been already obtained, the calculation of normalizing parameters in said basis vector calculating step and the execution of said feature extracting step are omitted, and said feature extracting step extracts the features using the basis vectors and the normalizing parameters that have been already obtained.

8. A text mining method for extracting features of documents as claimed in claim 1, wherein said basis vector calculating step comprises:

an initializing step of initializing a value of the basis vector;

a basis vector updating step of updating the value of the basis vector;

a variation degree calculating step of calculating a variation degree of the value of the basis vector;

a judging step of making a judgment whether a repetition process is to be terminated or not using the variation degree of the basis vector; and

a counting step of counting the number of times of said repetition process.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

9. A text mining method for extracting features of documents as claimed in claim 8, wherein said basis vector updating step updates the basis vector using a current value of the basis vector, the term-document matrix and an updating ratio controlling the updating degree of the basis vector.

10. A text mining method for extracting features of documents as claimed in claim 9, wherein, when all basis vectors and normalizing parameters required in extracting the features have been already obtained, the calculation of normalizing parameters in said basis vector calculating step and the execution of said feature extracting step are omitted, and said feature extracting step extracts the features using the basis vectors and the normalizing parameters that have been already obtained.

11. A text mining method for extracting features of documents as claimed in claim 8, wherein, when all basis vectors and normalizing parameters required in extracting the features have been already obtained, the calculation of normalizing parameters in said basis vector calculating step and the execution of said feature extracting step are omitted, and said feature extracting step extracts the features using the basis vectors and the normalizing parameters that have been already obtained.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

12. A text mining method for extracting features of documents as claimed in claim 1, wherein, when all basis vectors and normalizing parameters required in extracting the features have been already obtained, the calculation of normalizing parameters in said basis vector calculating step and the execution of said feature extracting step are omitted, and said feature extracting step extracts the features using the basis vectors and the normalizing parameters that have been already obtained.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

13. A text mining apparatus for extracting features of documents using a term-document matrix consisting of vectors corresponding to index terms representing the contents of the document, wherein contributions of the index terms act on respective elements of the term-document matrix, said apparatus comprising:

basis vector calculating means for calculating a basis vector spanning a feature space, in which mutually associated documents and terms are located in proximity with each other, based on a steepest descent method minimizing a cost;

feature extracting means for calculating a parameter for normalizing the features using the term-document matrix and the basis vector and extracting the features on the basis of the parameter; and

term-document matrix updating means for updating the term-document matrix to a difference between the term-document matrix, to which the basis vector is not applied, and the term-document matrix, to which the basis vector is applied.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

14. A text mining apparatus for extracting features of documents as claimed in claim 13, wherein the cost is defined as a second-order cost of the difference between the term-document matrix, to which the basis vector is not applied, and the term-document matrix, to which the basis vector is applied.

15. A text mining apparatus for extracting features of documents as claimed in claim 14, wherein said basis vector calculating means comprises:

initializing means for initializing a value of the basis vector;

basis vector updating means for updating the value of the basis vector;

variation degree calculating means for calculating a variation degree of the value of the basis vector;

judging means for making a judgment whether a repetition process is to be terminated or not using the variation degree of the basis vector; and

counting means for counting the number of times of said repetition process.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

16. A text mining apparatus for extracting features of documents as claimed in claim 15, wherein said basis vector updating means updates the basis vector using a current value of the basis vector, the term-document matrix and an updating ratio controlling the updating degree of the basis vector.

17. A text mining apparatus for extracting features of documents as claimed in claim 16, wherein, when all the basis vectors and normalizing parameters required in extracting the feature have been already obtained, the calculation of normalizing parameters by said basis vector calculating means and the execution of said feature extracting means are omitted, and said feature extracting means extracts the features using the basis vectors and the normalizing parameters that have been already obtained.

18. A text mining apparatus for extracting features of documents as claimed in claim 15, wherein, when all basis vectors and normalizing parameters required in extracting the features have been already obtained, the calculation of normalizing parameters by said basis vector calculating means and the execution of said feature extracting means are omitted, and said feature extracting means extracts the features using the basis vectors and the normalizing parameters that have been already obtained.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

19. A text mining apparatus for extracting features of documents as claimed in claim 14, wherein, when all basis vectors and normalizing parameters required in extracting the features have been already obtained, the calculation of normalizing parameters by said basis vector calculating means and the execution of said feature extracting means are omitted, and said feature extracting means extracts the features using the basis vectors and the normalizing parameters that have been already obtained.

20. A text mining apparatus for extracting features of documents as claimed in claim 13, wherein said basis vector calculating means comprises:

initializing means for initializing a value of the basis vector;

basis vector updating means for updating the value of the basis vector;

variation degree calculating means for calculating a variation degree of the value of the basis vector;

judging means for making a judgment whether a repetition process is to be terminated or not using the variation degree of the basis vector; and

counting means for counting the number of times of said repetition process.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

21. A text mining apparatus for extracting features of documents as claimed in claim 20, wherein said basis vector updating means updates the basis vector using a current value of the basis vector, the term-document matrix and an updating ratio controlling the updating degree of the basis vector.

22. A text mining apparatus for extracting features of documents in text mining as claimed in claim 21, wherein, when all basis vectors and normalizing parameters required in extracting the feature have been already obtained, the calculation of normalizing parameters by said basis vector calculating means and the execution of said feature extracting means are omitted, and said feature extracting means extracts the features using the basis vectors and the normalizing parameters that have been already obtained.

23. A text mining apparatus for extracting features of documents as claimed in claim 20, wherein, when all basis vectors and normalizing parameters required in extracting the features have been already obtained, the calculation of normalizing parameters by said basis vector calculating means and the execution of said feature extracting means are omitted, and said feature extracting means extracts the features using the basis vectors and the normalizing parameters that have been already obtained.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

24. A text mining apparatus for extracting features of documents as claimed in claim 13, wherein, when all basis vectors and normalizing parameters required in extracting the features have been already obtained, the calculation of normalizing parameters by said basis vector calculating means and the execution of said feature extracting means are omitted, and said feature extracting means extracts the features using the basis vectors and the normalizing parameters that have been already obtained.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

25. A computer program product for being executed in a text mining apparatus for extracting features of documents using a term-document matrix consisting of vectors corresponding to index terms representing the contents of the documents, wherein contributions of the index terms act on respective elements of the term-document matrix, the computer program product comprising:

basis vector calculating step of calculating a basis vector spanning a feature space, in which mutually associated documents and terms are located in proximity with each other, based on a steepest descent method minimizing a cost;

feature extracting step of calculating a parameter for normalizing the features using the term-document matrix and the basis vector and extracting the features on the basis of the parameter; and

term-document matrix updating step of updating the term-document matrix to a difference between the term-document matrix, to which the basis vector is not applied, and the term-document matrix, to which the basis vector is applied.

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111